

University of Groningen

Unintended consequences and the case of abduction. Reply to Roberto Festa

Kuipers, Theo A.F.

Published in:
EPRINTS-BOOK-TITLE

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Kuipers, T. A. F. (2005). Unintended consequences and the case of abduction. Reply to Roberto Festa. In *EPRINTS-BOOK-TITLE* University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Theo A. F. Kuipers

**UNINTENDED CONSEQUENCES AND THE CASE OF ABDUCTION
REPLY TO ROBERTO FESTA**

In his essay Roberto Festa elaborates the issues of (dis)similarities and possible interactions between logic and philosophy of science put forward by Johan van Benthem, in 1982, primarily in view of ICR, but also in view of SiS. Hence, the essay provides a natural transition from Volume 1, which mainly deals with topics of ICR, to Volume 2, which mainly deals with topics of SiS. In particular, as Festa makes clear, for the comparison of logic and philosophy of science it is plausible to emphasize the dominant research perspective in both ICR and SiS, viz. explicative research. The nature of this type of research is only explicitly thematized in SiS (pp. 8, 9, 18, 58-9, 248, 263-4). Festa is certainly right in stressing that the (plausible) definition of progress in explicative research presupposes the falsifiability of provisional explications, while the specific argumentation that an evident example or non-example is treated properly or that a condition of adequacy is, or is not, fulfilled may amount to proving a more or less deep theorem, or characterizing a more or less interesting countermodel. These points illustrate the claim in ICR (p. 130-1) and SiS (p. 248) that (explicative) research in philosophy should satisfy, *mutatis mutandis*, the three general principles of testability (PT), separate evaluation (PSE) and improvement (PI), where the latter presupposes the principle of comparative evaluation (PCE). Hence, Festa's essay pleases me very much. In this reply I would like to concentrate on the role of "unintended consequences," neglected by Festa, in speaking of progress in the empirical sciences, logic, and philosophy (of science). Moreover, I shall illustrate it by the recent finding of the straightforward abductive nature of Inference to the Best Theory (IBT), as opposed to the standard conception of Inference to the Best Explanation (IBE).

In: R. Festa, A. Aliseda and J. Peijnenburg (eds.), *Confirmation, Empirical Progress, and Truth Approximation (Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 83)*, pp. 521-525. Amsterdam/New York, NY: Rodopi, 2005.

Unintended Consequences in the Empirical and Non-Empirical Sciences

Although Festa refers to several paragraphs and passages in SiS dealing with concept explication, including the definition of progress in explicative research (SiS, p. 264), surprisingly enough he neglects my remark about unintended explications:

However, it is also considered to be very important that the proposed explication turns out to give rise to unintended explications, that is, to satisfactory explications of related concepts and intuitions. This type of success is the analogue of the extra, i.e., predictive or external, success of explanatory programs. Again the question is whether this form of success is formally defensible as a necessary condition for progress, but the fact remains that in practice this type of explicative success plays an important role. (SiS, p. 18)

In ICR, Subsection 7.5.1, I deal extensively with the last mentioned question regarding explanatory programs in view of the basic version of the structuralist theory of truthlikeness and conclude:

In sum, ad hoc repair of a theory will seldom be a real improvement without unexpected extra success. In other words, comparative HD-evaluation of an ad hoc repair will either lead to unexpected extra successes of the new theory or extra successes of the old theory that could have been, but were not, explicitly expected before. Hence, besides some qualifications, the intuitions of Popper and Lakatos with respect to ad hoc repairs and novel facts are largely justified. [However, I]nstead of a ban on ad hoc changes, they can be allowed, provided they are subjected to comparative HD-[evaluation] with the original theory. (ICR, pp. 168-9)¹

Hence, the first question is whether there is a similar story to tell about ad hoc improvements of explications. Unfortunately, the detailed argumentation in Subsection 7.5.1 referred to leans heavily on the unknown but fixed character of the postulated target, the truth, that is, the set of nomic possibilities determined by the domain and the vocabulary. In contrast – although less extreme than in the case of the material version of explicative research, that is, design research (SiS, pp. 282-3) – explicative research is guided by a more or less known target which may or may not be changed. Hence, the prospects for an analogous way of reasoning are not impressive.

However this may be, the intuition stated in the first quote remains: an explication of a concept generating related but unintended explications is more impressive than an explication that does not. Fortunately, close reading of this plausible specification of the intuition suggests that the above quoted comparison with ad hoc repairs of empirical theories may not be adequate. In the case of ad hoc repairs we are talking about a given theory and a revised version that solves a given problem of the former. In the case of unintended

¹ Note that I inserted 'However' at the beginning of the last sentence in order to stress the deviating moral. Moreover, I replaced 'HD testing' by 'HD evaluation', since using the former expression here must have been a slip of the falsificationist pen.

explications we are primarily comparing two explications of the same concept in terms of their side effects. Hence, suppose we have two explications of a given concept that are equally successful in the sense suggested by my definition of explicative progress (SiS, p. 264), quoted by Festa. Suppose, moreover, that the one generates one or more explications of, apparently, related concepts and the other does not: as far as we know, of course. In other words, the one leads to conceptual unification whereas the other does not. In several contributions and replies in both volumes there have been hints about non-empirical merits of empirical theories that may be taken into account in addition to empirical ones. Here we seem to come across a variant of this type of merits of concept explication in addition to the straightforward merits in terms of evident examples and non-examples and conditions of adequacy. However, in the empirical case it was, at least sometimes, possible to argue that non-empirical merits are functional for truth approximation in an indirect way, viz. in the case of a certain type of aesthetic criteria (Kuipers 2002). In the present case of conceptual unification such an argument seems less likely, roughly for the same reasons as we have met before: the target is more or less known and changeable. However, if we take the set of all concepts that become related by an explication of the concept we started with, we can of course apply a generalized version of the definition of explicative progress. Then we will find that the resulting set of explications of the set of concepts unified in this sense, starting from an explication of a given concept, is better than an alternative explication of this given concept that is equally successful with respect to this concept but has nothing to offer with respect to the concepts that are related to it by the first explication. However, to get a fair comparison one has to bring in the best available, presumably isolated, explications of these concepts. Even if all the resultant one-by-one comparisons show equal success, we are inclined to say that the unified explication of a set of concepts is superior to the sum total of disconnected explications of them. However, it is again difficult to see how this could be functional for truth approximation. Hence, unification remains a non-empirical merit of concept explication.

So far we have assumed that one explication of a given concept was compared with another existing explication. However, the unifying merit of explications of concepts generating unintended explications may be generalized in the sense that such explications are better than any (already existing or not yet existing) explication that is equally successful with respect to an initial concept but does not generate related explications.

Note that a similar story can be told about theoretical unification. If one theory has links with other theories whereas an (existing or hypothetical) empirically equally successful competitor remains isolated, this certainly is

considered to be a merit of the former. David Atkinson argues in his contribution to the companion volume that the main merit of string theory is this kind of theoretical unification. He even suggests that the research program underlying string theory may have to be seen as exemplifying a type of program that differs from the four standard types characterized in SiS, that is, descriptive, explanatory, design and explicative programs. A research program may merely aim at unification of (domains and) theories, that is, aim at a unification that seems to have, for practical reasons, no prospects of empirical evaluation.

Inference to the Best Theory as a Case of Abduction

The general merit of conceptual unification suggested above certainly applies to the structuralist explication of the notion of truthlikeness presented in detail in ICR, relative to equally successful but isolated explications of this notion and of its relatives as generated by this explication. In ICR I summarize its merits in this respect as follows:

- ... the structuralist theory of truthlikeness has generated a number of unexpected and unintended results. The most important ones are:
 - a justification of fundamental methodological rules, in particular the rule to choose the most successful theory (Subsection 7.3.3.),
 - an explanation of the success of the natural sciences in terms of truth approximation (Subsection 7.3.3.),
 - an explanation and justification of the non-falsificationist behavior of scientists (Section 6.2., 6.3. and 7.3.),
 - a corrective explication of so-called ‘inference to the best explanation’ (Subsection 7.5.3, [called “inference to the best theory,” see also Kuipers 2004]),
 - an explication of Popper’s bad luck, i.e., the convincing failure of his at first sight very plausible definition of truthlikeness (Section 8.1.),
 - an explication of the correspondence theory of truth as an intralevel intuition (Section 8.2.),
 - ... an explication of dialectical concepts [Section 8.3] (ICR, p. 198).

Since the appearance of ICR in 2000, two other unintended explications can be added to this list, viz.

- an explication of the intuition of many scientists and philosophers that aesthetic criteria are indirectly functional for empirical progress and even truth approximation (Kuipers 2002),
- an explication of the intuition of several philosophers that ‘inference to the best explanation’ is a kind of abduction in the sense of Peirce (Kuipers 2004).

The last mentioned explicative success is of a special kind. As included in the first list, the structuralist explication of truthlikeness generated a corrective explication of the idea of the “inference to the best explanation” (IBE), leading to “inference to the best theory” (IBT). The last success may also be seen as a proof of the claim that the generated explication of IBE, viz. IBT, satisfies the plausible condition of adequacy that an explication of IBE should make it an abductive rule of inference in the paradigmatic sense of Peirce.

I would like to give a rough idea of the basic versions of these related explications. In ICR I argued that IBE should not be explicated as “inference to the best unfalsified theory, if any, as true,” as is usually suggested, but as “inference to the best theory, if any, whether falsified or not, as the closest to the truth,” i.e., as IBT. In my forthcoming paper I argue that the latter, but not the former, satisfies the crucial second premise of Peirce’s scheme for abduction (Peirce 1958, p. 189):

The surprising fact *C* is observed.
But if *A* were true, *C* would be a matter of course.
Hence, there is reason to suspect that *A* is true.

That is, the hypothesis of being closer to the truth² makes a matter of course of the, as such, surprising fact of remaining empirically more successful,³ viz. due to the Success Theorem, see Festa’s paper. In contrast, the hypothesis of being true does not do so. This may be an unintended merit of the structuralist explication of the notion of truthlikeness that can be particularly appreciated by logicians interested in abductive logic. However this may be, it at least illustrates once again that relatively simple theorems may nevertheless be very important because of their far-reaching consequences with respect to the unification of concepts and intuitions.

REFERENCES

- Kuipers, T. (2002). Beauty, a Road to The Truth. *Synthese* **131** (3), 291-328.
- Kuipers, T. (2004). Inference to the Best Theory, Rather Than Inference to the Best Explanation. Kinds of Induction and Abduction. In: F. Stadler (ed.), *Induction and Deduction in the Sciences*, pp. 25-51. Dordrecht: Kluwer Academic Publishers.
- Peirce, C.S. (1958). *Collected Papers of Charles Sanders Peirce*, vol. 5. Cambridge, MA: Harvard University Press.

² More precisely, interpreting *A* as the hypothesis that the one theory is closer to the truth than the other.

³ More precisely, interpreting *C* as the (surprising) comparative fact that the first theory, besides having an extra success, is empirically at least as successful as the second, and remains so after further evaluation.